ECAI 2024
U. Endriss et al. (Eds.)

doi:10.3233/FAIA240827

MCAHNN: Multi-Channel EEG Emotion Recognition Using Attention Mechanism Based on Householder Reflection

Qinglong Liu[1], Wenhao Jiang[1], Shihang Ding[1], Kaixuan Wang[1], Hongjian Bo[2], Cong Xu[1], Lin Ma[1] and Haifeng Li[1,*]

[1]Faculty of Computing Harbin Institute of Technology Harbin, China
[2]Shenzhen Academy of Aerospace Technology Shenzhen, China

Abstract. Emotions are integral to human cognition, exerting a profound influence on physiological responses, cognitive processes, and decision-making capabilities. Electroencephalography (EEG)-based emotion classification provides a significant methodological approach for the exploration of emotional states. Despite its potential, most current methodologies face challenges in delineating the representational patterns across different brain regions and in effectively classifying emotions from EEG signals. In response, a novel model for emotion recognition is proposed in this paper, which utilizes a multi-channel attention mechanism, designated as MCAHNN. This model incorporates Householder Reflection to enhance the attention mechanism, facilitating the extraction of inter-channel EEG features and simulating inter-regional brain dynamics. Furthermore, 1D convolution is employed to analyze intra-channel relationships. The proposed model has been evaluated on the publicly available DEAP dataset and further tested on the SEED dataset. Experimental results confirm that the MCAHNN model achieves state-of-the-art performance, demonstrating its effectiveness in classifying emotions within multi-center datasets.Code is publicly available at https://github.com/Oreoreoreor/MCAHNN.

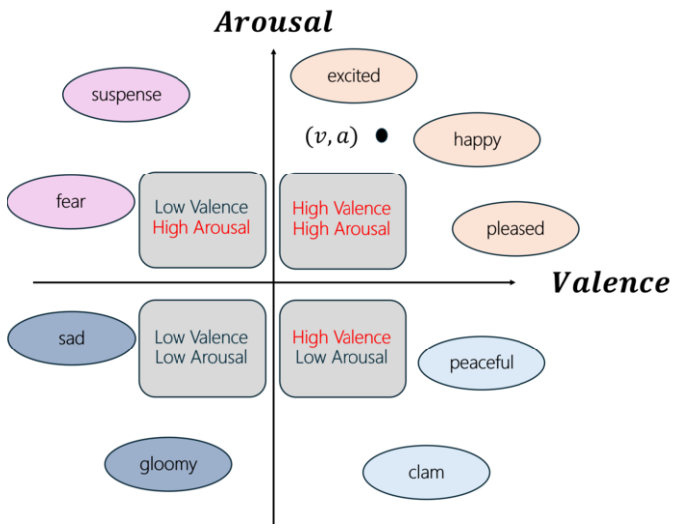


Figure 1. Russel's emotion model.

1 Introduction

Emotion is a fundamental component of human cognition, characterized by sophisticated neural processes. The generation of emotions entails synchronized synaptic activities within designated brain regions[18, 19], resulting in electrical signals transmitted to the scalp surface. Moreover, various emotions elicit distinct patterns of activity across specific brain regions, highlighting the complexity of emotional processing. Emotion is frequently assessed utilizing multidimensional scales, including arousal and valence[20] (as shown in Fig 1), underscoring the neural foundations of emotional responses and their discernment through neurophysiological assessments[1].

Electroencephalography[3] (EEG) is a non-invasive electrophysiological technique that records brain electrical activity. In comparison to alternative brain signal detection techniques, EEG is notable for its objectivity, convenience, accessibility, and enhanced temporal resolution. Such features facilitate the precise and objective monitoring of variations in emotional states [23]. Significant advances have been made in emotion recognition using EEG signals across various public datasets. Among these, the DEAP [13] dataset is widely employed for EEG-based emotion recognition, favored for its high data quality and effective emotion measurement methods. Researchers[4] typically use the frequency-domain feature of EEG or the time-frequency domain feature as a biomarker to measure emotion. For example, Lu et al. [29] used short-time Fourier transform to extract five frequency bands from EEG signals as frequency domain features, and introduced the concept of entropy to calculate the frequency band uncertainty as EEG features, and Varun Bajaj et al. [2] used a variety of wavelet bases to perform wavelet transform on EEG signals to extract the time-frequency features of the data, and put the features into a multi-classification least squares support vector machine (MC-LSSVM) for classification. Additionally, the deployment of deep learning techniques enables the extraction of emotion-related patterns from EEG signals, potentially serving as valuable biomarkers. For example, wavelet-transformed spectra of raw EEG signals were utilized as inputs to a 2D convolutional neural network by Kwon, Yea-Hoon et al. [14], enhancing the EEG feature extraction.

Despite these advancements, a standardized set of EEG biomark-

* Corresponding Author. Email: lihaifeng@hit.edu.cn

ers for emotions, derived from frequency domain or time-frequency analyses, has yet to be established. Furthermore, most prior methods overlooked the relationships among different EEG channels. Emotional responses in the brain are closely linked to interactions between various brain regions[17], which may delineate distinct emotional categories. Therefore, studying the relationships among different EEG electrodes could shed light on the representational patterns of various brain regions during emotional experiences, thereby enhancing the model's ability to discriminate emotions.

Currently, Yongqiang Yin et al. [27] devised an emotion recognition approach based on EEG channel relationships (ERDL), utilizing Graph Convolutional Networks[7] (GCN) to analyze inter-channel connections. The adjacency matrix of the GCN was initialized using brain functional connectivity among 32 electrodes from the DEAP dataset, yielding binary classification accuracies of 90.45% for Valence and 90.60% for Arousal. Tengfei Song et al. [22] developed a multi-channel EEG emotion recognition method using dynamic graph convolutional neural networks (DGCNN). The adjacency matrix was initialized using the geometric distances between EEG channels and subsequently trained within the network to determine effective channel relationships, achieving a tri-classification accuracy of 90.4% on the SEED dataset. Yue Gao et al. [10] proposed an EEG-GCN model that employs feature-adaptive channel selection, where inter-channel distances, processed through ReLU and Softmax functions, were used to populate the graph convolution adjacency matrix. This model achieved binary classification accuracies of 81.77% for Valence and 81.95% for Arousal on the DEAP dataset and a tri-classification accuracy of 85.65% on the SEED dataset. Xuefen Lin et al. [16] proposed an CSGNN model that calculates the Phase Lag Index (PLI) between different channels to initialize the adjacency matrix for the Graph Attention Networks[26] (GAT), which updates the matrix based on attention scores from the network, reaching a four-class classification accuracy of 91.00% on the DEAP dataset and 90.22% on the SEED dataset.

Although the methods mentioned have attained high recognition accuracies on public datasets such as DEAP, they exhibit limitations including weak flexibility and a strong dependence on prior knowledge. Furthermore, these approaches encounter challenges in fully exploring the latent relationships within the data and in investigating the correlations between different channels and emotions.

In this work, we introduce a deep learning network termed the Multi-Channel Attention based on Householder Reflection Neural Network (MCAHNN). The MCAHNN model is tailored to process EEG signals in emotion recognition tasks, incorporating a 1D-CNN to extract time-frequency features from EEG signals and an attention mechanism based on Householder Reflection to minimize inter-channel information loss. Notably, the MCAHNN excels in both sentiment classification and in-depth exploration of inter-channel EEG features. The model comprises three main modules: **(1) Feature embedding module**, consisting of a shallow CNN network that extracts time-frequency features and maps the raw EEG data into a feature space enriched with emotion-discriminative information; **(2) Attention-based encoder module**, which leverages an attention mechanism framework and incorporates Householder Reflection to project embeddings into an orthogonal space for efficient channel selection; and **(3) Emotion decoder module**, tasked with predicting emotional class labels by mapping features to the label space.

To verify the applicability of the model, we conducted training, testing, and analysis on the DEAP dataset, supplemented by additional testing on the SEED [28, 8] dataset. The experimental results demonstrate that the proposed method achieves accuracies of 94.96%, 94.98%, and 93.43% on the valence binary classification task, the arousal binary classification task, and the valence-arousal quadruple classification task on DEAP, respectively. Moreover, it achieved an overall classification accuracy of 93.36% on the SEED dataset.

The main contributions of this paper are as threefold:

i) A channel-wise self-attention-based emotion recognition model is introduced, which reduces the inter-channel loss and improves recognition accuracy by adaptively learning the functional topology between EEG channels.

ii) An EEG channel integration method using Householder Reflection is proposed to narrow the search space of the self-attention mechanism, reducing training costs.

iii) Experimental results demonstrate the model's capability in discerning channel relationships on multi-center datasets.

2 Method

In this section, we present the overall structure and implementation details of the proposed MCAHNN. The architecture of MCAHNN is illustrated in Fig 2. The model consists primarily of three modules: the feature embedding module, the encoder module, and the classifier. The feature embedding module is utilized to extract time-frequency features from channel-wise EEG signals. The encoder module dynamically learns the spatial relationships of the embedded features and perform integration to minimize feature loss across channels. Subsequently, a linear classifier is employed to utilize the spatiotemporal features, where the cross-entropy is adopted as the loss function to calculate the discrepancy between the true and predicted values.

2.1 Feature embedding

A four-layer convolutional neural network is carefully tailored to embed the raw EEG signals into time-frequency features. To be specific, the first two layers are designed to extract the time-frequency features of the EEG, while the last two layers focus on extracting the frequency domain features. Each convolutional layer is composed of a 1D-CNN and a LeakyReLU activation function. The parameters for the convolutional kernel size, output channels, and stride for the four layers are respectively: (1×128, 128, stride 1), (1×32, 128, stride 1), (1×32, 98, stride 1), and (1×32, 1, stride 1), as demonstrated in Table 1.

Table 1. Parameters of the convolution in feature embedding module.

Layer	Kernel size	Out channel	Stride
conv1	(1×128)	128	1
conv2	(1×32)	128	1
conv3	(1×32)	98	1
conv4	(1×32)	1	1

Let the original input signal be $\{x_1, \cdots, x_i, x_{i+1}, \cdots, x_C\} \in \mathbb{R}^{C\times T}$ where C denotes the number of channels of the EEG signal, and T denotes the length of each sequence, then we have:

$$Conv_{block\,i} = leaky\_Relu\left(conv_{1d_i}\left(x_i\right)\right) \quad i = 1, 2, 3, 4 \tag{1}$$

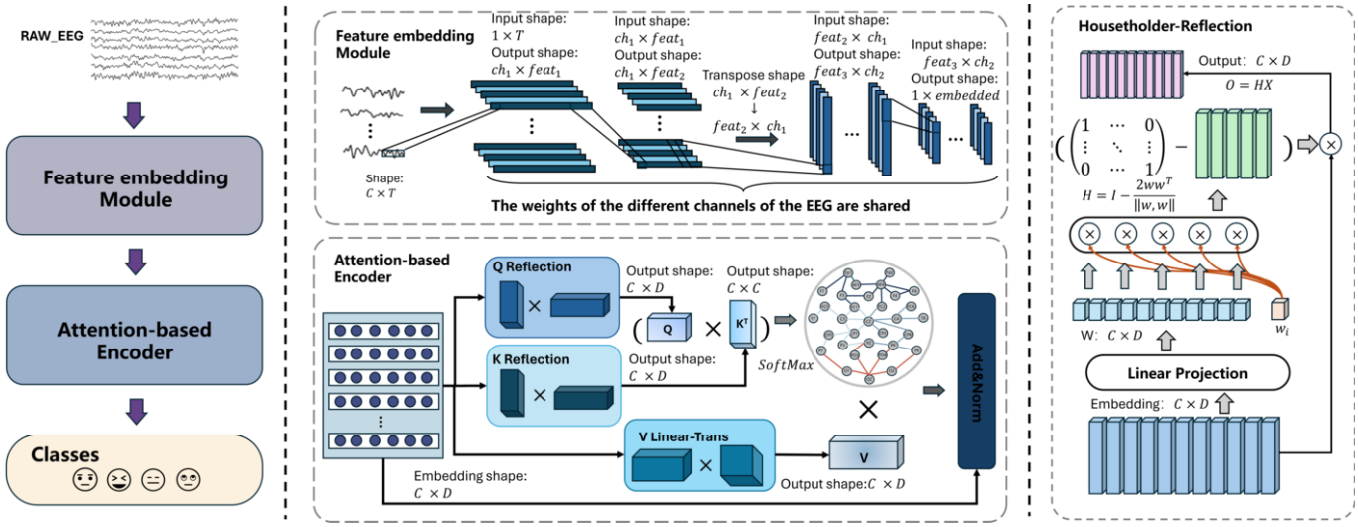


Figure 2. Architecture of the MCAHNN. The left figure shows the framework of the model, which is composed of three main parts. The structure of the feature embedding module and the attention-based encoder module is shown in the middle figure, the structure of attention based on Householder Reflection is shown on the right figure .

We take $x_1, x_2 \cdots, x_C$. one by one from X as inputs to this module to get the embedding of each channel:

$$x_{embedding_i} = Conv\_module\,(x_i) \quad i = 1, 2, \cdots, C \tag{2}$$

In Eq.(2), a single embedding module is shared across all C channels, with shared weights. This approach allows for the features to be extracted independently within and across channels.

2.2 Encoder

In constructing the encoder, we incorporate Householder Reflection into the Self-Attention mechanism to design the Encoder module. This process involves learning optimal reflection planes that allow the embeddings to be precisely mapped into desired subspaces. By doing so, it enhances the ability of the Self-Attention mechanism to focus and project the embeddings more accurately within the neural network architecture.

Let the Encoder input be:

$$X_e = \{x_{embedding_1}, \cdots, x_{embedding_i}, \cdots, x_{embedding_c}\} \tag{3}$$

The output obtained after the Encoder is:

$$X_c = Encoder\,(X_e), \quad X_c \in \mathbb{R}^{C \times D} \tag{4}$$

where C denotes the number of channels of the EEG signal, D denotes the size of the embedding, X_e is the output of the embedding of the original signal in the previous subsection, and X_c is the output of the inter-channel features after fusion.

Householder Reflection, was initially proposed by A.C. Aitken[24]. Later, Alston S. Householder[12] highlighted its significance in linear algebra. This transformation is an orthogonal transformation used to describe reflection across a hyperplane. The key principle behind a Householder transformation is to construct an orthogonal matrix that, when multiplied by a vector, results in a reflection of that vector across a hyperplane. Its capability to mirror vectors precisely across hyperplanes makes it invaluable in algorithms that require orthogonal transformations or reflections, enhancing both the accuracy and efficiency of these computational processes.

Let w be the unit vector, then the Householder Reflection matrix is:

$$H = I - 2ww^T \tag{5}$$

For w, with $w^T w = 1$, then:

$$Hw = \left(I - 2ww^T\right) w = w - 2w\left(w^T w\right) = -w \tag{6}$$

For any vector v perpendicular to w, with $w^T v = 0$, then:

$$Hv = \left(I - 2ww^T\right) v = v - 2w\left(w^T v\right) = v \tag{7}$$

For the vector w itself, the Householder Reflection reverses it, and for the vector v perpendicular to w, the Householder Transformer has no effect on it. Therefore, for any vector α, after Householder Reflection reverses the components of α that are parallel to w, and the components that are perpendicular to w remain unchanged. As a result, $H\alpha$ is the specular reflection of α in a hyperplane with w as normal vector.

Householder Reflection-based Attention is proposed to construct orthogonal matrices, achieving different mirror reflections of embeddings to derive the key (K) and query (Q) matrices. This approach aligns features across different channels using a minimal number of parameters, thereby reducing the search space for attention scores. Compared to traditional Self-Attention, this method significantly decreases training costs by leveraging the efficiency of Householder transformations to streamline the alignment and computation process within the attention mechanism.

The computation of the conventional self-attention[25] mechanism is defined as follows:

$$Attention\,(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right) V \tag{8}$$

In the above equation, Q, K, and value (V) matrix are obtained from the input features X after linear transformation, which represent the query, key and value vectors of the sample, respectively. The values are calculated by $Q = XW_q$, $K = XW_k$, $V = XW_v$, where $X \in \mathbb{R}^{C\times D}$, $W_q \in \mathbb{R}^{D\times d_k}$, $W_k \in \mathbb{R}^{D\times d_k}$, $W_v \in \mathbb{R}^{D\times d_k}$, W_q, W_k, W_v are the transformation matrices of the corresponding vectors, and d_k is the dimension of the feature.

The Householder Transform is introduced during the mapping of X to Q and K. Q is computed as follows:

$$Q = \left(I - 2\frac{uu^T}{\|u,u\|}\right)X \tag{9}$$

where $u = XW_p$, $W_p \in \mathbb{R}^{D\times 1}$, $u \in \mathbb{R}^{C\times 1}$, W_p is a projection matrix. Thus $I - 2\frac{uu^T}{\|u,u\|}$ is a Householder Reflection orthogonal matrix generated by X. From Eq.(9), we can generate the corresponding mirror surface to reflect to get Q. The full expression for QK^T is as follows:

$$QK^T = \left(I - 2\frac{XW_{pQ}\left(XW_{pQ}\right)^T}{\|XW_{pQ}, XW_{pQ}\|}\right)X \left(\left(I - 2\frac{XW_{pK}\left(XW_{pK}\right)^T}{\|XW_{pK}, XW_{pK}\|}\right)X\right)^T \tag{10}$$

Substituting Eq.(5), into Eq.(10) yields:

$$HA(X) = softmax\left(\frac{\left(H_Q X\right)\left(H_K X\right)^T}{\sqrt{d_k}}\right)XW_V \tag{11}$$

Let QK^T denote the inter-channel similarity relation have:

$$Channel\_Weight = Softmax\left(QK^T\right) \tag{12}$$

Therefore, the feature fusion between channels is expressed as:

$$HA(X) = Channel\_WeightXW_v \tag{13}$$

At the same time each channel should maintain the original information X so the fused channel information should be:

$$HA(X) + X \tag{14}$$

Therefore, after Eq.(13) is brought into layer normalization, the whole Encoder calculation process is as follows:

$$Encoder = Layer\_normal\left(HA(X) + X\right) \tag{15}$$

2.3 Classifier

The classifier of the model consists of a fully connected layer, where the classifier utilize gets the features from the Encoder and splices them by channel to achieve a mapping from the feature space to the category space, and subsequently performs a Softmax computation on the results of the mapping to convert the mapping into a probability distribution for the category:

$$\hat{y} = Softmax\left(Linear\left(concat\left(x_1, \cdots x_i, \cdots x_C\right)\right)\right) \quad i = 1, 2, \cdots, C \tag{16}$$

where $concat\left(x_1, \cdots x_i, \cdots x_C\right) \in \mathbb{R}^{N\times(C\times D)}$, $O \in \mathbb{R}^{N\times k}$, N denotes the number of samples and k denotes the number of categories.

Moreover, the Cross-Entropy is utilized as loss function to calculate the distance between the true and predicted samples.

$$Loss\left(\hat{y}, y\right) = -\sum_{i=1}^{k}\hat{y}_i \log y_i \tag{17}$$

3 Experiments

In this section, we will utilize the DEAP dataset, which is extensively used for EEG emotion recognition, to validate the effectiveness of our method. Additionally, we will employ the SEED test to verify the robustness of our model.

3.1 Data and Processing

In this subsection, we detail the datasets used for our experiments (as outlined in Table 2). We describe the preprocessing steps applied to the data and the methodology used for dividing the dataset labels.

3.1.1 Dataset

The DEAP dataset, which integrates physiological and visual signals, is extensively employed for emotion classification and analysis. It comprises EEG and physiological signals collected from 32 healthy participants (16 males and 16 females). Participants viewed 40 one-minute videos while EEG data were recorded using a 32-lead Biosemi ActiveTwo device, configured according to the '10-20' international standard (as shown in Figure 3a). Beyond the 32 EEG channels, the dataset includes recordings from 16 additional physiological channels covering ophthalmological and cardiological signals. After viewing each video, subjects rated the videos on Valence, Arousal, and Dominance on a scale from 1 to 9 to assess emotional impact.

Table 2. Information on the DEAP dataset and the SEED dataset.

Items	DEAP	SEED
Subjects	32	15
Stimulate	Videos	Movie clips
Videos	40	15
Videos duration	60s	about 4 minutes
Channels	40	62
Data	40×32×40×8064	15×15×62×5×235
Label	40×32×1	15×15×1

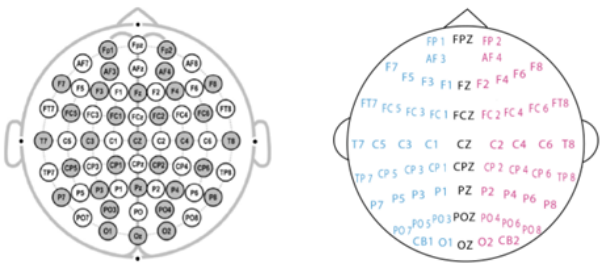
The SEED dataset, provided by the BCMI laboratory, consists of EEG recordings from 15 subjects (7 males and 8 females) who viewed film clips eliciting positive, neutral, and negative emotions. Each clip lasted between 3 to 5 minutes. The data collection was conducted over three separate sessions for each participant, spaced approximately one week apart. EEG signals were captured using a 62-channel ESI NeuroScan device (as depicted in Figure 3b) at a sampling rate of 1 kHz.

3.1.2 Pre-Processing

In the DEAP dataset, the raw signal acquired by each subject was downsampled to 128 Hz. Channels such as EOG were excluded, and the data were filtered with a 4-45 Hz band-pass filter. Additionally, eight electrodes unrelated to EEG signals were removed. This process isolated the data comprising the 3s pre-viewing resting state, the 60s of data during video viewing, and another 3s of resting data post-viewing. The initial 3s resting data served as the baseline for the EEG signals. The preprocessed signal was formatted as 32 (subjects) × 40 (trials) × 32 (channels) × 7680 (samples). Subsequently, the processed 60s EEG signal was segmented into 30 non-overlapping 2s

Table 3. EEG emotion recognition Results Comparison. Comparison of the mean accuracy (acc) and standard deviation (std) of the MCAHNN with other baseline models on the SEED three-classification task, the DEAP two-classification task, and the four-classification task, respectively.

Method	DEAP			SEED
	Valence	Arousal	V-A	Pos/Neg/Neu
	acc/std(%)	acc/std(%)	acc/std(%)	acc/std(%)
SVM	63.09/6.22	69.65/9.41	56.02/8.61	72.88/6.58
DBN	73.67/7.54	78.04/6.18	70.28/7.62	76.03/7.56
CNN	74.52/6.09	78.08/6.23	71.33/6.81	76.91/7.72
ERDL[27]	90.45/7.12	90.60/4.15	83.13/6.93	83.35/8.26
DGCNN[22]	89.23/3.85	90.57/8.47	88.07/5.27	90.40/8.49
EEG-GCN[10]	81.77/5.58	81.95/7.71	80.63/5.47	84.65/7.49
CSGNN[16]	91.96/5.84	92.59/6.09	91.00/5.70	90.22/3.67
MCHANN	**94.96/0.89**	**94.98/0.99**	**93.50/1.38**	**93.36/1.49**



(a) DEAP (b) SEED

Figure 3. Electrode location of EEG devices. (a) represents the channel locations of the EEG acquisition devices used for the DEAP dataset, (b) represents the channel locations of the EEG acquisition devices used for the SEED dataset

sequences (as shown in Figure 4) for testing. The segmented format was 32 (subjects) × 40 (trials) × 32 (channels) × 30 × 256 (samples). A dimensional transformation of this segmented data resulted in a training dataset format of 38400 × 32 × 256.

Figure 4. Data Segmentation.

The Valence and Arousal dimensions were selected for labeling in quadratic/quadruple classification tasks. For the binary classification on each dimension, labels were assigned by dividing the ratings such that scores greater than or equal to 5 were labeled as 1, indicating high Valence or Arousal, and scores less than 5 were labeled as 0, indicating low Valence or Arousal, represented as $0, 1$.

In the quadratic classification task involving both Valence and Arousal, the categories were defined as follows: low Valence with low Arousal was set to 0, low Valence with high Arousal was set to 1, high Valence with low Arousal was set to 2, and high Valence with high Arousal was set to 3. These categories are denoted as $0, 1, 2, 3$.

In the SEED dataset, for comparison purposes, we utilize the processed Differential Entropy (DE) features as model inputs. The data format for each subject is 62 (channels) × 5 × 235 (samples), where '5' represents the five frequency bands: δ, θ, α, β, and γ, each containing 235 features. The labels for the triple categorization in SEED—negative, neutral, and positive—are mapped as $-1, 0, 1$, respectively.

3.2 Training and Result

In this subsection, we detail the training configuration of the proposed MCAHNN model, the outcomes of comparative and ablation experiments, and an analysis of the model's capabilities along with activation maps.

3.2.1 Model Training

During training, to fully leverage the dataset and mitigate evaluation bias from arbitrary data partitioning, we employ 10-fold cross-validation. This strategy minimizes the model's reliance on a specific training set and enhances the assessment of its predictive performance.

Table 4. Configuration during model training.

Model configuration	Value or Type
Batch size	128
Dropout	0.2
Optimizer	Adam
Learning rate	0.001
β_1, β_2	0.9,0.99
Activation function	Leaky ReLU
Pooling method	Max pooling

The training configuration is displayed in Table 4, where the batch size is set at 128, and the model typically converges after approximately 70 iterations. We use SoftMax cross-entropy as the loss function and Adam as the optimizer with a learning rate of 0.001; the decay rates for the first-order and second-order moments are set at 0.9 and 0.99, respectively. Figure 5 illustrates the training performance on the DEAP dataset for the Valence-Arousal quadruple classification. The model begins to stabilize around iteration 20 and fully converges by iteration 70, achieving a final accuracy of 93.50%.

3.2.2 Classification Results

In order to investigate the ability of the model to perform emotion recognition, the recall, precision and F1-score of MCAHNN under different classification tasks on the DEAP dataset are computed.

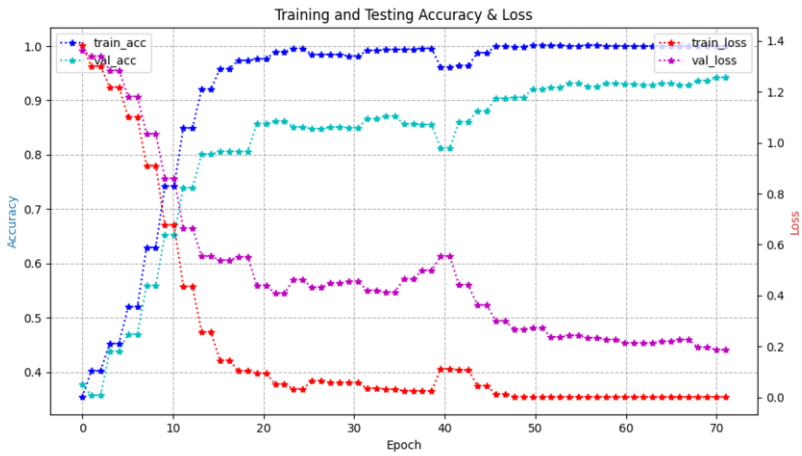


Figure 5. Accuracy and Loss for EEG emotion recognition. The blue and green lines indicate the change in accuracy between the training and validation sets during training, and the red and purple lines indicate the change in loss between the training and validation sets, respectively.

Table 5. Evaluation indicators of the MCAHNN in DEAP.

Label	Evaluation indicators		
	Recall	Precision	F1-score
LV	92.89%	94.63%	0.9375
HV	96.07%	94.81%	0.9544
LA	92.79%	94.40%	0.9358
HA	96.16%	95.04%	0.9559
LVLA	91.18%	93.60%	0.9237
LVHA	91.48%	93.27%	0.9235
HVLA	93.04%	93.93%	0.9348
HVHA	95.44%	92.8%	0.9410

Table 5 illustrates the capability of MCAHNN in classifying Valence and Arousal labels. The evaluation metrics indicate that the MCAHNN method performs robustly across all emotion types, notably excelling in recognizing High Valence/Arousal with an F1-score exceeding 0.95. In the quadruple categorization task, the recognition of the High Valence High Arousal label is particularly effective, achieving an F1-score of 0.94, which is higher than those of the other three emotion categories. The F1-scores for the four emotion labels in this task are 0.9237, 0.9235, 0.9348, and 0.9410, respectively. It is observed that emotions associated with lower Valence are more prone to misclassification compared to those with Arousal, whereas higher Valence is recognized more accurately. This trend suggests that emotions with higher Valence exhibit more distinct and recognizable patterns, leading to more reliable detection.

3.2.3 Comparative Analysis

To assess the performance of our proposed MCAHNN model, we conducted comparative analyses with several representative models, categorized into three groups: machine learning methods such as Support Vector Machines (SVMs)[5]; deep learning methods including Deep Belief Networks (DBNs)[11], Convolutional Neural Networks (CNNs); and deep learning approaches based on EEG channels like ERDL, DGCNN, EEG-GCN, and CSGNN. The comparative performance across different classification tasks in the DEAP and SEED datasets is detailed in Table 3.

On the SEED dataset, the proposed MCAHNN model achieved an accuracy of 93.36%. As for the DEAP dataset, it reached 94.96% accuracy in the Valence and Arousal binary classification tasks, and 93.50% in the Valence-Arousal quadruple classification task. Notably, in the DEAP Valence-Arousal quadruple classification, the MCAHNN outperformed other models significantly, improving accuracy by 37.48%, 23.22%, 22.17%, 10.37%, 5.43%, 12.87%, and 2.50% respectively, against various models. These results suggest that: **(1)** The MCAHNN model efficiently extracts emotional patterns by focusing on inter-regional brain relationships, enhancing emotion identification in EEG signals compared to non-EEG channel-based models; and **(2)** its advanced Householder-reflection-based attention mechanism effectively fuses features across channels, reducing inter-channel loss, providing substantial benefits over other EEG channel-based models. Furthermore, the MCAHNN model consistently outperforms other models in both datasets, demonstrating low variability and excellent stability. The test results underscore the versatility and robustness of our proposed method, affirming its efficacy in classifying emotions across different environments.

3.2.4 Ablation Study

To further explore the capabilities of Attention-based Householder Reflection in MCAHNN for extracting inter-channel features, we conducted two ablation experiments within the Valence-Arousal quadruple classification task on the DEAP dataset. These experiments involved omitting the Encoder module and using only traditional self-Attention. The results, presented in Table 6, demonstrate that our method enhances accuracy by 3.38% over traditional self-Attention. Additionally, the incorporation of Attention-based Householder Reflection significantly improves the F1-score. This improvement indicates that the method more effectively captures the relationships between channels and integrates features across channels. Consequently, the model exhibits heightened sensitivity to variations in emotional states, enhancing its overall emotion recognition capabilities.

Table 6. EEG emotion recognition Results in Ablation Study. Recall, precision, and F1-score of MCAHNN for different labels under different classification tasks on DEAP dataset.

Method	Accuracy	Precision	Recall	F1-Score
Without Attention	73.09%	72.51%	71.89%	0.7219
With Self-Attention	90.12%	89.44%	89.40%	0.8941
MCAHNN	**93.50%**	**93.40%**	**92.78%**	**0.9307**

3.2.5 Activation maps

As shown by Fig 6, the $Channel\_Weight$ matrix in the MCAHNN Encoder module is visualized, and the learned parameters represent the correlation of brain regions represented by different channels with emotion. The redder color indicates the higher the activation of the brain region, while the bluer the color indicates the lower the activation of the brain region.

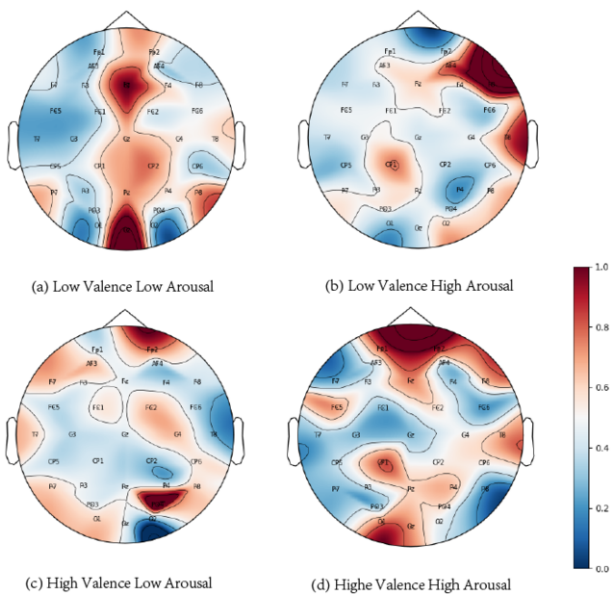


Figure 6. Activation maps of MCAHNN under different emotions. The activation map of the brain regions of the subjects under different emotions learned by MCAHNN under the DEAP four-classification task.

In the DEAP dataset analysis using the MCAHNN model, distinct brain activation patterns were observed correlating with specific emotional stimuli. Activation predominantly occurred in the middle frontal and middle occipital lobes when subjects were exposed to stimuli with low valence and low arousal (LVLA). In contrast, stimuli characterized by low valence and high arousal (LVHA) elicited significant activity in the right frontal lobe and the right temporal lobe. For high valence and low arousal (HVLA) conditions, the primary activation was noted in the right parts of the frontal and occipital lobes. Lastly, stimuli with high valence and high arousal (HAHV) predominantly activated the frontal lobe and the right occipital lobe. These findings indicate substantial variations in the brain regions activated by each emotional category—LVLA, LVHA, HVLA, and HAHV—highlighting distinct neural representation patterns corresponding to various human emotions. This supports the hypothesis that different emotions engage specific neural circuits, as referenced in studies [6, 21, 15].

4 Conclusion

In this paper, we introduce an emotion recognition model utilizing a multi-channel attention mechanism that integrates features from convolutional embedding and attention mechanisms, and incorporates Householder Reflection to minimize inter-channel loss and enhance the accuracy of emotion classification. This is achieved by learning the functional topology between EEG channels. Comparative analyses with existing models across various emotion classification tasks demonstrated that our model, MCAHNN, effectively simulates the architecture of different brain regions involved in emotional processing. It excels in extracting latent features from EEG emotion signals, thereby significantly improving emotion recognition accuracy. The robust performance of the MCAHNN model on the various datasets underscores its efficacy in complex dataset conditions and highlights its versatility and robustness. Additionally, ablation studies confirm that integrating Householder Reflection into the attention mechanism markedly improves the model's capability to extract channel features and significantly boosts emotion recognition accuracy. Furthermore, we visualized the $Channel\_Weight$ matrix derived through MCAHNN, revealing distinct neural representation patterns under various emotional states. These findings contribute valuable insights into the neural correlates of emotional processing[9].

Acknowledgments

This study is supported by the National Natural Science Foundation of China (U20A20383), Basic and Applied Basic Research of Guangdong (2021B1515120052), Independent Research Exploration Projects of Songjiang Laboratory(Grant No.SL20230203).

References

[1] R. Adolphs. How should neuroscience study emotions? by distinguishing emotion states, concepts, and experiences. *Social cognitive and affective neuroscience*, 12, 10 2016. doi: 10.1093/scan/nsw153.

[2] V. Bajaj and R. B. Pachori. *Detection of Human Emotions Using Features Based on the Multiwavelet Transform of EEG Signals*, pages 215–240. Springer International Publishing, Cham, 2015. ISBN 978-3-319-10978-7. doi: 10.1007/978-3-319-10978-7_8. URL https://doi.org/10.1007/978-3-319-10978-7_8.

[3] A. Biasiucci, B. Franceschiello, and M. M. Murray. Electroencephalography. *Curr Biol*, 29(3):R80–R85, Feb 2019.

[4] M. X. Cohen. Where does eeg come from and what does it mean? *Trends in Neurosciences*, 40(4):208–218, 2017. ISSN 0166-2236. doi: https://doi.org/10.1016/j.tins.2017.02.004. URL https://www.sciencedirect.com/science/article/pii/S0166223617300243.

[5] C. Cortes and V. N. Vapnik. Support-vector networks. *Machine Learning*, 20:273–297, 1995. URL https://api.semanticscholar.org/CorpusID:52874011.

[6] R. J. Davidson and N. A. Fox. Asymmetrical brain activity discriminates between positive and negative affective stimuli in human infants. *Science*, 218(4578):1235–1237, 1982. doi: 10.1126/science.7146906. URL https://www.science.org/doi/abs/10.1126/science.7146906.

[7] M. Defferrard, X. Bresson, and P. Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS'16, page 3844–3852, Red Hook, NY, USA, 2016. Curran Associates Inc. ISBN 9781510838819.

[8] R.-N. Duan, J.-Y. Zhu, and B.-L. Lu. Differential entropy feature for eeg-based emotion classification. In *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*, pages 81–84, 2013. doi: 10.1109/NER.2013.6695876.

[9] J. Fei, T. Wang, J. Zhang, Z. He, C. Wang, and F. Zheng. Transferable decoding with visual entities for zero-shot image captioning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3136–3146, October 2023.

[10] Y. Gao, X. Fu, T. Ouyang, and Y. Wang. Eeg-gcn: Spatio-temporal and self-adaptive graph convolutional networks for single and multi-view

eeg-based emotion recognition. *IEEE Signal Processing Letters*, 29: 1574–1578, 2022. doi: 10.1109/LSP.2022.3179946.

[11] G. E. Hinton. Deep belief networks. *Scholarpedia*, 4:5947, 2009. URL https://api.semanticscholar.org/CorpusID:7905652.

[12] A. S. Householder. Unitary triangularization of a nonsymmetric matrix. *J. ACM*, 5(4):339–342, oct 1958. ISSN 0004-5411. doi: 10.1145/320941.320947. URL https://doi.org/10.1145/320941.320947.

[13] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras. Deap: A database for emotion analysis ;using physiological signals. *IEEE Transactions on Affective Computing*, 3(1):18–31, 2012. doi: 10.1109/T-AFFC.2011.15.

[14] Y.-H. Kwon, S.-B. Shin, and S.-D. Kim. Electroencephalography based fusion two-dimensional (2d)-convolution neural networks (cnn) model for emotion recognition system. *Sensors*, 18(5), 2018. ISSN 1424-8220. doi: 10.3390/s18051383. URL https://www.mdpi.com/1424-8220/18/5/1383.

[15] Y. Li, L. Wang, W. Zheng, Y. Zong, L. Qi, Z. Cui, T. Zhang, and T. Song. A novel bi-hemispheric discrepancy model for eeg emotion recognition. *IEEE Transactions on Cognitive and Developmental Systems*, 13 (2):354–367, 2021. doi: 10.1109/TCDS.2020.2999337.

[16] X. Lin, J. Chen, W. Ma, W. Tang, and Y. Wang. Eeg emotion recognition using improved graph neural network with channel selection. *Computer Methods and Programs in Biomedicine*, 231:107380, 2023. ISSN 0169-2607. doi: https://doi.org/10.1016/j.cmpb.2023.107380. URL https://www.sciencedirect.com/science/article/pii/S0169260723000470.

[17] K. A. Lindquist, T. D. Wager, H. Kober, E. Bliss-Moreau, and L. F. Barrett. The brain basis of emotion: a meta-analytic review. *Behav Brain Sci*, 35(3):121–143, Jun 2012.

[18] M. Malezieux, A. S. Klein, and N. Gogolla. Neural circuits for emotion. *Annual review of neuroscience*, 46:211–231, 7 2023. ISSN 0147-006X. doi: 10.1146/annurev-neuro-111020-103314. [Online; accessed 2024-04-22].

[19] G. Prete, P. Croce, F. Zappasodi, L. Tommasi, and P. Capotosto. Exploring brain activity for positive and negative emotions by means of eeg microstates. *Scientific Reports*, 12(1):3404, Mar 2022. ISSN 2045-2322. doi: 10.1038/s41598-022-07403-0. URL https://doi.org/10.1038/s41598-022-07403-0.

[20] J. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39:1161–1178, 12 1980. doi: 10.1037/h0077714.

[21] E. K. Silberman and H. Weingartner. Hemispheric lateralization of functions related to emotion. *Brain and Cognition*, 5(3): 322–353, 1986. ISSN 0278-2626. doi: https://doi.org/10.1016/0278-2626(86)90035-7. URL https://www.sciencedirect.com/science/article/pii/0278262686900357.

[22] T. Song, W. Zheng, P. Song, and Z. Cui. Eeg emotion recognition using dynamical graph convolutional neural networks. *IEEE Transactions on Affective Computing*, 11(3):532–541, 2020. doi: 10.1109/TAFFC.2018.2817622.

[23] O. Sourina and Y. Liu. Eeg-enabled affective applications. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, pages 707–708, 2013. doi: 10.1109/ACII.2013.125.

[24] H. W. Turnbull and A. C. Aitken. An introduction to the theory of canonical matrices. *Nature*, 130:867–867, 1932. URL https://api.semanticscholar.org/CorpusID:29811510.

[25] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 6000–6010, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 9781510860964.

[26] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio. Graph attention networks. In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=rJXMpikCZ.

[27] Y. Yin, X. Zheng, B. Hu, Y. Zhang, and X. Cui. Eeg emotion recognition using fusion model of graph convolutional neural networks and lstm. *Applied Soft Computing*, 100:106954, 2021. ISSN 1568-4946. doi: https://doi.org/10.1016/j.asoc.2020.106954. URL https://www.sciencedirect.com/science/article/pii/S1568494620308929.

[28] W.-L. Zheng and B.-L. Lu. Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks. *IEEE Transactions on Autonomous Mental Development*, 7(3): 162–175, 2015. doi: 10.1109/TAMD.2015.2431497.

[29] W.-L. Zheng, J.-Y. Zhu, and B.-L. Lu. Identifying stable patterns over time for emotion recognition from eeg. *IEEE Transactions on Affective Computing*, 10:417–429, 2016. URL https://api.semanticscholar.org/CorpusID:15982451.